

Image-to-image translation of label-free molecular vibrational images for a histopathological review using the UNet+/seg-cGAN model

YUNJIE HE,¹  JIASONG LI,¹ STEVEN SHEN,² KAI LIU,¹ KELVIN K. WONG,^{1,3} TIANCHENG HE,¹ AND STEPHEN T. C. WONG^{1,2,3,*}

¹Translational Biophotonics Laboratory, Systems Medicine and Bioengineering Department, Houston Methodist Cancer Center, Houston, USA

²Pathology and Genome Medicine Department, Houston Methodist Hospital, Weill Cornell Medicine, Houston, USA

³T.T. and W. F. Chao Center for BRAIN, Houston Methodist Academic Institute, USA

*st Wong@houstonmethodist.org

Abstract: Translating images generated by label-free microscopy imaging, such as Coherent Anti-Stokes Raman Scattering (CARS), into more familiar clinical presentations of histopathological images will help the adoption of real-time, spectrally resolved label-free imaging in clinical diagnosis. Generative adversarial networks (GAN) have made great progress in image generation and translation, but have been criticized for lacking precision. In particular, GAN has often misinterpreted image information and identified incorrect content categories during image translation of microscopy scans. To alleviate this problem, we developed a new Pix2pix GAN model that simultaneously learns classifying contents in the images from a segmentation dataset during the image translation training. Our model integrates UNet+ with seg-cGAN, conditional generative adversarial networks with partial regularization of segmentation. Technical innovations of the UNet+/seg-cGAN model include: (1) replacing UNet with UNet+ as the Pix2pix cGAN's generator to enhance pattern extraction and richness of the gradient, and (2) applying the partial regularization strategy to train a part of the generator network as the segmentation sub-model on a separate segmentation dataset, thus enabling the model to identify correct content categories during image translation. The quality of histopathological-like images generated based on label-free CARS images has been improved significantly.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Coherent anti-Stokes Raman scattering (CARS) is a nonlinear four-wave mixing process that is used to enhance the weak (spontaneous) Raman signal by several orders [1]. CARS is a label-free, 3-dimensional (3D), real-time optical imaging modality that can provide biological tissue images with sub-cellular level resolution [2–4]. The CARS modality has been used in disease research and pathological diagnosis by imaging target tissue both *ex vivo* and *in vivo* [5–8]. In this manuscript, we are developing a method to convert the images from a high-resolution, label-free, 3-D capable optical imaging system to H&E pseudo-stained imaging. Only a few imaging modalities fall into this category, such as autofluorescence imaging, second-harmonic generation (SHG), and third-harmonic generation (THG). These imaging modalities can achieve high resolution label-free images, but they are not suitable for all tissue types [9]. CARS is sensitive to certain chemical bonds and is not limited to a particular tissue type. Therefore, by tuning the wavelength of the laser source in our CARS system, we can image an array of different tissue types. The chemical bond contrast has the potential for pathological diagnosis similar to histopathological stained images, such as through hematoxylin-eosin (H&E) staining, to view the

tissue architecture. However, pathologists and clinical investigators are used to interpreting the images in H&E presentation.

Prior research shows the potential for stain-less staining or virtual staining [10,11]. Stainless staining for computed histopathology is possible by translating infrared (IR) microscopy to H&E images, without the need of labeling agents. It can provide high quality images in some cases. However, IR microscopy is constrained in its applications and outcomes due to its fundamental limitations, including the high spectral similitude between spectra of different bacteria, high signal overlap and widening of adjacent signals, incompatibility for the acquisition of spectra into samples with content of intrinsic water, background signals of substrate producing “false” signals not associated with the sample, signals with weak intensity resulting in a low signal-to-noise ratio (SNR), and excessive spectral noise [12,13]. In addition, tissue autofluorescence is defined as a natural ability of tissue to fluoresce when exposed to a certain light wavelength and is not suitable for imaging most tissue types [14].

In this paper, we developed a deep learning enabled image-to-image translation model that can map label-free CARS images into an H&E image presentation automatically to sidestep the need for additional tissue preparation and staining steps. Generative adversarial networks (GAN) were introduced by Ian J. Goodfellow et al. in 2014 [15–19]. Their brand-new structure gained lots of attention in various areas such as data augmentation, data enhancement, and data generation. Particularly in the computer vision domain, GAN are more flexible and provide a new angle for image translation and augmentation. GAN are composed of two parts: the generator and the discriminator. The generator produces images from random noise (or from input images with prior information) to confuse the discriminator, while the discriminator aims to classify the input image as a pseudo but realistic looking image. The generator and discriminator compete with each other during training and achieve optimized states while loss functions converge. GAN are accompanied by fully connected layers in the beginning and then become more powerful after combination with convolutional layers.

Owing to GAN’s effectiveness, many mutated structures have been developed. For example, a conditional GAN indicates that the generator input can be either random noise or other meaningful data input, paving the way for image translation [20]. CycleGAN can be used to train an image-to-image translation model, which is not highly dependent on paired datasets, by its additional inverse mapping structure [21]. WGAN introduces Wasserstein distance to improve the converging process during training [22,23]. LSGAN and hinge loss based GAN bring different objective functions to ameliorate the performance [24–28]. Further, abundant innovations related to GAN have been created for specific tasks, such as Δ -GAN using semi-supervised learning, and DiscoGAN and DualGAN applied to transfer learning [29–31]. GAN usually have better performance for image translation than traditional methods. The goals of image translation vary across medical imaging applications, and thus it is hard to find an objective function for a specific task and even harder to generalize it. Nevertheless, the discriminator in a GAN overcomes these problems by being trained along with the generator. This optimization process allows a GAN to avoid creating a complex loss/dice function that is often not generalizable.

The reliability of output images is always the key issue in applying GAN in medical image generation and translation, and many improvements have been made. For example, the Laplacian generative adversarial networks (LAPGAN) were designed by Denton et al. to upgrade the resolution of output images [32]. Isola et al. introduced Pix2pix, a supervised image translation model [18], and the appearance of a Pix2pix GAN has improved the performance significantly. The L1 regularization loss in Pix2pix improved the robustness over other conditional GAN, particularly for medical image generation, translation, and even denoising. Chen et al. applied GAN in magnetic resonance imaging (MRI) reconstruction and computed tomography (CT) generation based on MRI [27]. Medical image translation is a rapidly emerging research area of

GAN [17,33–35], notably their applications to image translation between CT and MRI scans [21,36–38].

In this paper, the subject is translating CARS images to H&E style by using GAN [39]. If the CARS-H&E translation technology is successful, pathological diagnosis based on a CARS 3D capable image modality has the potential to replace H&E staining and allow label-free imaging of tissue presentation in familiar H&E format but without the need for staining in practical use. However, there are hardly any image-to-image translation applications reported on CARS-H&E stained images. Recent works in this area are related to translating H&E style images based on multimodal CARS/TPEF/SHG [40,41]. They showed great potential for translating multichannel label-free microscopy to pseudo-H&E staining microscopy. Unfortunately, the resulting image contrast is not usually accurate, containing misleading information for pathological diagnosis.

In the current study, we propose a more accurate and robust image-translation method and illustrate the model using thyroid cancer tissue images acquired by CARS microscopy. In contrast to prior work that transferred the color style of H&E images, we leverage the ground truth tissue segmentation labels in H&E microscopy during model training and use it to significantly improve the accuracy in image translation to pseudo-H&E presentation where tissue labels are not available. Our model integrates UNet+ with seg-cGAN, conditional generative adversarial networks with partial regularization of segmentation. It replaces UNet with UNet+ as the Pix2pix cGAN's generator to enhance pattern extraction and richness of the gradient and applies the partial regularization strategy to train a part of the generator network as the segmentation sub-model on a separate segmentation dataset, thus enabling the model to identify correct content categories during image translation.

2. Materials and methods

The automated image-to-image translation pipeline to map label-free CARS images to pathologic H&E images by integrating image preprocessing and translation with an improved Pix2Pix conditional GAN model, called UNet+/seg-cGAN is shown in Fig. 1.

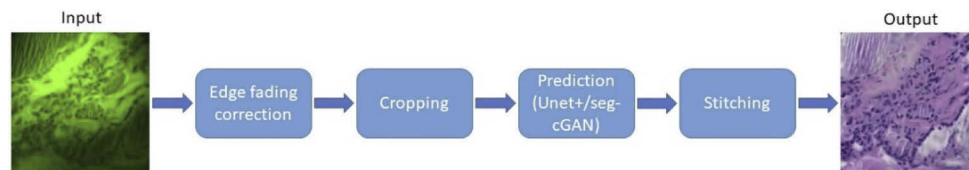


Fig. 1. The automated imaging pipeline developed for CARS-H&E image translation.

2.1. Data collection

The following two datasets were established for model training purposes: (1) CARS-H&E image translation dataset (200 scans/10 cases), and (2) CARS segmentation dataset (200 scans/10 cases). The two datasets are paired and share identical CARS images scanned from thyroid tissues. To construct the CARS-H&E dataset, the same group of tissues was H&E stained and photographed. Furthermore, CARS images were labeled by pathologists and saved as mask images. Therefore, we obtained the CARS segmentation dataset by combining these masks with its corresponding CARS images.

Human thyroid tissue was obtained from patients undergoing thyroid/parathyroid surgery at Houston Methodist Hospital, Houston, Texas, USA and Shanghai General Hospital, Shanghai, P.R. China, following an institutional review board approval. The excised tissue samples were cut into 5 mm chunks and then immediately snap-frozen in liquid nitrogen for storage. Frozen tissue samples were passively thawed for 30 minutes at room temperature before CARS imaging.

The sample was first imaged by CARS system and immediately sent for H&E staining afterward. (Fig. 2)

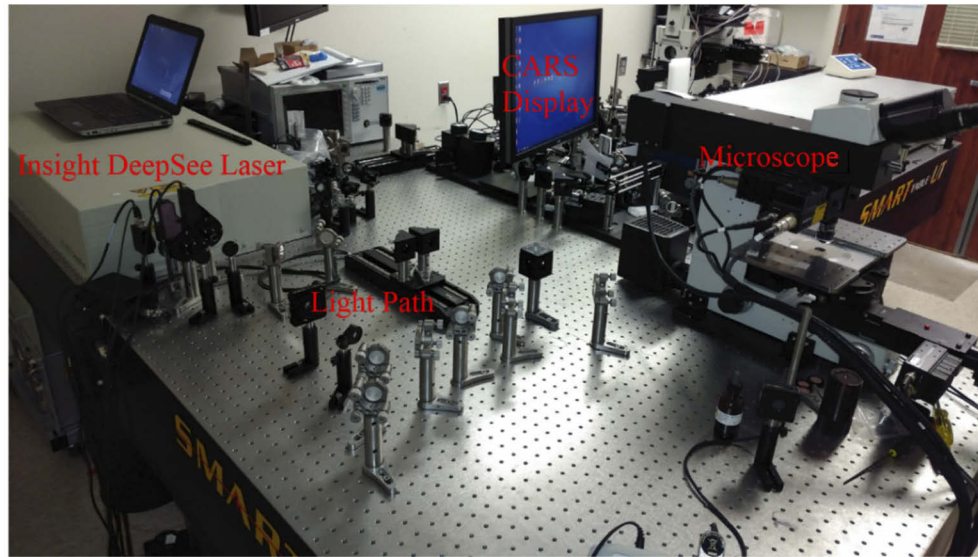


Fig. 2. The imaging system for acquiring data.

The tissue samples were placed on a 170 μm cover slide (VWR, Radnor, PA, USA) and then inverted on an imaging chamber to avoid possible compression. For Artificial intelligence-augmented CARS (iCARS) imaging, the pump beam was tuned to 802 nm while the Stokes beam was fixed at 1040 nm to probe the symmetric stretching frequency of CH₂ bond at 2845 cm^{-1} [42]. iCARS signals were generated at 663 nm. The image (1024 \times 1024 pixels) was acquired and displayed in real-time using the ThorImage 3.0 software (Thorlabs, Inc), and the average laser output power was about 75 mW for pump beam and 35 mW for Stokes beam. Bright-field images of the H&E slides were examined with an Olympus BX51 microscope as a standard control.

2.2. Data preprocessing methods

It is important to preprocess raw data before the model training. The data preprocessing improves the training with high probability in practice. The following three sub-sections describe the key preprocessing procedures that we applied to optimize the model performance.

2.2.1. Edge fading correction

Microscopy images generated by CARS often have fading issues near or at the edges of the images. This margin seriously affects the statistical distribution of image pixel values and reduces the model performance if it is uncorrected. To resolve fading issues, we deployed BaSiC method proposed by Tiangying et al., a tool that applies intensity normalization based on low rank and sparse decomposition to the target images [43]. This largely reduced the edge fading compared to the original CARS images.

2.2.2. Image registration

Many GAN, such as style-transfer GAN, allow training without paired image couples [44–46]. However, these GAN can rarely be applied to medical image translation due to their lack of precision. Moreover, matched image pairs will always bring a huge improvement to the performance of the model. Since each tissue is imaged twice on different scanners, the two

images will be slightly misaligned or distorted even if the tissue was photographed carefully. Image registration is indispensable for application to impaired images for optimizing the dataset quality.

Image registration includes two procedures, location of reference points and image transformation [47,48]. Finding the reference points is one of the most important steps in image registration, because no matter how fancy the transforming method is, the output will be totally wrong if the reference points were not located correctly. Nevertheless, the imaging principles are vastly different between CARS and H&E images. CARS images only have one channel while H&E images are in the RGB format. Consequently, the automated registration methods easily make mistakes on locating reference points, thus ruining the entire registration operation.

We implemented a semi-automatic image registration Python tool to display the high probability reference points, inspired by [49]. The tool allows manual adjustment of the reference points when the points are incorrectly located. A registration sample is shown in Fig. 3(B). After having verified the reference points, the affine transformation was applied to obtain the registered H&E image (see the third image in Fig. 3(B)).

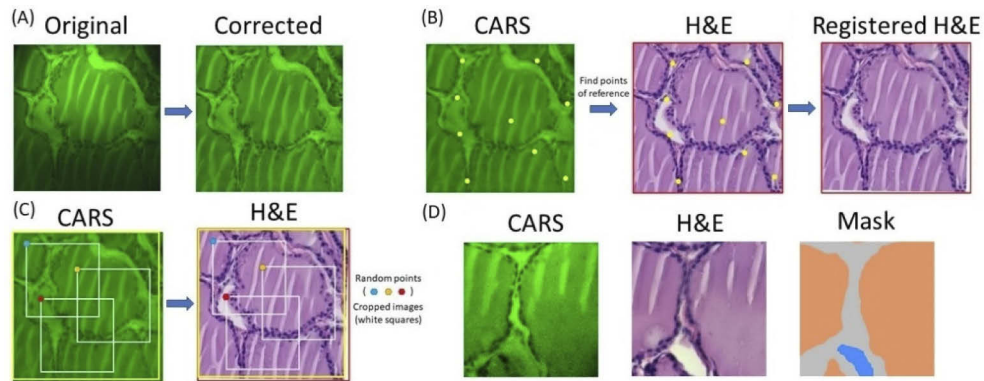


Fig. 3. (A) representative sample of edge shade correction (B) The image registration pairs CARS and H&E images generated from the same tissue sample. The reference points (yellow dots) were used for transformation to pair images. The image in the last column is the H&E image after registration. (C) The figure shows how the augmentation process works. The yellow rectangles are the cropped useful areas of original images after registration. The three points with different colors are random points generated from uniform distribution. The white squares in the images are the cropped images. (D) Cropped CARS and H&E images of the same thyroid tissue and corresponding mask image which are ready for the model training.

2.2.3. Image augmentation

After registering the original H&E image, we can further augment data by cropping and rotating image pairs and adding noise to the CARS images. The cropping size is set as 256×256 .

$$\text{uniform}([0, s - 256])$$

The cropping strategy uses the uniform distribution to pick up the left-up corner of the augmented image and crop it off with the size of 256×256 . (Figure 3(C))

The 200 raw image pairs were acquired from scanners and were randomly split into training (100), validation (40), and testing (60) datasets. The scans from same case were not located in the different datasets. After registration and augmentation, the number of image pairs were

multiplied by 40 (i.e. train:validation:test = 4000:1600:2400),

$$f(x) = \begin{cases} -1, & \text{if } x + z < 0 \\ 1, & \text{if } x + z > 255 \\ \frac{x+z}{127.5} - 1, & \text{if } 0 \leq x + z \leq 255 \end{cases}$$

where f is the data generator function, x is the pixel values of the original input images, and z is the added noise with $z \sim \text{uniform}([-0.1, 0.1])$.

2.2.4. CARS segmentation dataset

The CARS segmentation dataset is based on the CARS-H&E image dataset. The CARS images were manually labeled by pathologists in three different types of content: thyroid follicle (orange), background (blue), and cells (grey). The segmentation dataset was simultaneously augmented along with the CARS-H&E image dataset. Finally, we have 4000, 1600, and 2400 image pairs in the training, validation, and test datasets (Fig. 3(D)).

2.3. Post-processing

Since a CARS image is cropped to 256×256 pieces before being entered into the model, we need to apply a stitching strategy to reconstruct the pseudo-H&E image from the outputs. Thus, we developed a special stitching strategy to efficiently orGANize the outputs. Since the model easily makes mistakes at the edge of the image, the input CARS image of the model was deliberately cropped partially overlapped between the pieces.

During the experiments, we found that the efficiency decreases if the overlapped area is too large, especially evident when the width of the overlapped edge is larger than half the size of the input image. However, the advantages of the stitching strategy will not be significant while the overlapped edges are extremely narrow. Moreover, our cropping strategy can find the optimized width capable of dividing the boundary length without remainder pixels. The following formula shows the way to find the optimized width of overlapped edges,

$$\min_{0 < x < \lfloor \frac{C}{2} \rfloor, x \in N} x \cdot 1_{\{(h-x) \bmod (C-x)=0\}}$$

where x is the optimized overlapped edge width, C is the cropped image size, and h represents the width of the original image size ($h = 256$ here).

Through the formula, we can decrease abrupt transition between adjacent cropped images, thereby minimizing loss of information, increasing precision, and smoothing the stitched edges. The equation ensures that the computational cost will not increase dramatically by finding a reasonable h .

The stitching method deals with the overlapped edges of outputs and determines the final image quality. Here we propose two efficient stitching dice.

2.3.1. Mean dice

The mean dice averages the pixel values from the overlapped area of outputs mapped from cropped CARS images. For the pixel p on the channel c , we have:

$$q_c = \frac{1}{M} \sum_{m \in \{1, 2, \dots, M\}} p_{cm}$$

where M is the number of overlapped images at the current pixel, p_{cm} is the predicted value of the pixel p at the channel c , and q_c represents the average value for the pixel p at channel c .

2.3.2. Median dice

The median dice calculates the pixel values' median from the overlapped area of outputs, which are mapped from cropped CARS images. For the pixel P on the channel c , we have:

$$q_c = \frac{1}{M} \text{Median}_{m \in \{1, 2, \dots, M\}} p_{cm}$$

where M is the number of overlapped images at the current pixel, p_{cm} is the predicted value of the pixel p at the channel c , and q_c is the median value for the pixel p at channel c .

2.3.3. Dynamic dice

Since the quality of a CARS image at the edge is far less than the middle of the image, the possibility of outliers will increase greatly. The median dice is thus a better choice at image edges because it is less sensible to the outliers. However, the mean dice is more accurate while the variance of pixel values is low. Therefore, the median dice is only applied when:

$$|\text{Mean}_{p_i} - \text{Median}_{p_i}| > \frac{1}{n} \sum_{j \in \{1, \dots, n\}} |\text{Mean}_{p_j} - \text{Median}_{p_j}|$$

where p_i is the current pixel, and n is the number of pixels in the target image. The dynamic dice will be efficient while the mean and median are vastly different, but this situation is rare.

2.4. Methods

Although the GAN model has brought huge benefits to image generation and transformation, there is still room for improvement in better presenting architectural details, especially when applying the model to highly sensitive patient images. We thus propose a new model, termed U-Net+&seg-GAN, to improve the fidelity and quality of medical image transformation. The new model significantly reduces the information loss during the CARS-H&E stained image translation and increases the accuracy compared to the prevalent Pix2pix conditional GAN model.

Considering that the quality of translated images from CARS is highly sensitive to information loss, the generator's structure has been carefully designed so that the information extracted by each convolutional layer can be efficiently utilized.

The image-to-image translation starts with a basic image reconstruction structure, encoder-decoder that has two desirable properties: (1) the encoder extracts the information in different levels by applying sequential blocks of neural network layers, and (2) the decoder can reconstruct the image in the same shape of the input [19,50]. The layers in the encoder-decoder architecture do not have to be convolutional. For example, digital images in MNIST dataset only have resolution of $28 \times 28 \times 1$ and are good to be raveled as input of the neural networks constructed by dense layers.

Unlike the basic encoder-decoder architecture, U-Net is most likely to be accompanied by convolutional layers, which is powerful for image feature extraction [51]. The reason is that the special skip-connection structure of U-Net allows the convolutional and deconvolutional layers to be directly connected while they are in the same block level. This design is useful to vary the gradients and acquire more information for the ending layers, hence improving the loss convergence process [52,53].

The Pix2pix GAN model uses the U-Net as its generator and achieves a better performance than the encoder-decoder. However, the limitation of U-Net remains: the skip-connection combinations between convolutional and deconvolutional blocks are restricted due to the same shape rule between convolutional block and its related deconvolutional block. To enrich the skip-connection and vary the gradients further, UNet+ is introduced.

2.4.1. UNet+ GAN

UNet+, originally designed for image segmentation, is introduced into the Pix2pix GAN model in our research. The UNet+ overcomes the limitations by its redesigned skip-connections. Figure 4(A) shows the UNet+, designed by Zhou et al., set as the GAN's generator. The skip-connections of U-Net were replaced by up and down samplings in different depths [50,54]. The downward, upward, and horizontal arrows in Fig. 4 represent the down-sampling, up-sampling, and skip-connections between convolutional blocks. Different depths in UNet+ vary the feature extraction levels and enrich more gradients during the loss optimization compared to the U-Net structure.

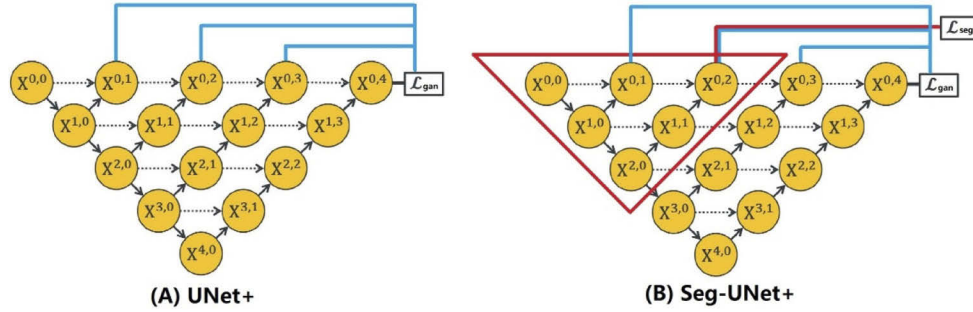


Fig. 4. (A) represents the structure of UNet+ GAN's generator. Layers $X^{0,1}$, $X^{0,2}$, and $X^{0,3}$ are directly concatenated and then connected to the final layer, which enriches the gradients of the loss function. (B) is the UNet+/seg-cGAN's generator. Comparing to (A), the UNet+/seg-cGAN's generator compiled the red triangle part as an image segmentation model. It was simultaneously trained based on the CARS segmentation dataset and with the entire GAN's generator network.

2.4.2. UNet+ seg-GAN

Although different categories of content have the same color gamut, they differ in biological expressions. For example, the white color may not indicate similar components in different content categories. As shown in Fig. 3(D), the white part with blue mask indicates the background while the thyroid follicle with white artifacts is labeled under orange mask. It is important to send to the generator a priori information, such as content categories, which will help the generator to make better translation decisions. The proposed UNet+/seg-cGAN method is designed to address this issue. Figure 4(B) shows its generator's structure. Based on UNet+, the part of networks in the red triangle was compiled with the categorical cross-entropy loss (L_{seg}) for the segmentation model, while the whole networks were compiled for cGAN's loss. Thus, the optimization problem becomes:

$$\min_G \max_D L_{cGAN}(D, G) = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] + \lambda L(G)$$

where G is the UNet+ generator, D is the convolutional discriminator, z represents noises generated from the uniform distribution added to CARS images during image augmentation, $L_{cGAN}(D, G)$ represents the general function, and x represents samples randomly chosen from the real H&E stained image dataset. The regularization part is multiplied by a constant λ (0.5-1), and θ represents the parameters related to segmentation:

$$L_1(G) = \mathbb{E}_{x,y,z} [|y - G(x, z)|]$$

$$L_2(G) = \mathbb{E}_{x,y,z} [|y - G(x, z)|^2]$$

$$L_{seg}(G_{\theta_{seg}}) = - \sum_{c=1}^C y_c \log G_{c,\theta_{seg}}(x, z)$$

$L_{seg}(G_{\theta_{seg}})$ represents the categorical loss of the segmentation model, where C is the number of the content categories in the segmentation dataset. Its related parameters θ are also involved in GAN's optimization. We name this strategy partial regularization, which restricts regularization effects in the specific layers and therefore injects additional information into the target model (see Algorithm 1).

Overall, we made the following improvements from Pix2pix cGAN. First, the UNet generator of Pix2pix cGAN was replaced by UNet+ to enhance the feature extraction during convolutional operations and to vary the loss function's gradients. Second, the front part of the generator was compiled as a segmentation sub-model. During the optimization of the whole UNet+, the sub-model was simultaneously trained on a segmentation dataset. All CARS images were labeled by mask of different content categories in this dataset. This strategy was named partial regularization.

3. Results

The models were trained on a TITAN RTX NVIDIA GPU with Adam optimizer (learning rate = $1e-3 \sim 1e-5$, beta1 = 0.6, and beta2 = 0.5). The training finished while the average change of generator's mean square error loss was less than $1e-4$.

Algorithm 1 The procedure for Training UNet+ seg-cGAN

Input: $(X_{CARS}, Y_{HE}, Y_{seg}) \sim P_{data}$ image pairs, number of epochs e , number of training batches n for each epoch

- 1: Model parameters randomly initialized $\theta \sim \mathcal{N}(0, \sigma^2)$
- 2: **for** $i \in \{1, \dots, e\}$ **do**
- 3: $(x_{cars}, y_{he}, y_{seg})_1, \dots, (x_{cars}, y_{he}, y_{seg})_n$ random sample batches
- 4: z_1, \dots, z_n from uniform distribution
- 5: **for** $j \in \{1, \dots, n\}$ **do**
- 6: Maximize GAN loss along θ_D :

$$\nabla_{\theta_D} (E_{x \sim P_{data}} [\log D(x)] + E_z [\log(1 - D(G(x)))])$$

- 7: Minimize GAN loss along θ_G :

$$\nabla_{\theta_G} (E_{x \sim P_{data}} [\log D(x)] + E_z [\log(1 - D(G(x)))]) + \lambda(L_1(G) + L_2(G))$$

- 8: Minimize segmentation loss along θ_{seg} sharing parameters with θ_G :

$$\nabla_{\theta_{seg}} L_{seg}(G_{\theta_{seg}})$$

- 9: end for
 - 10: end for
 - 11: **return** G
-

3.1. Model performance

In Fig. 5, three representative sample groups from the test dataset illustrate the fidelity and quality of pseudo H&E images. Owing to the differences of photographic mechanism between CARS and H&E imaging methods, many tissue details in the real H&E stained images cannot be found

in its corresponding CARS image. However, the UNet+ generator upgraded the utilization of information in CARS images and accordingly can maintain more biological structures during the image translation. Also, the sub-segmentation model can help identify the content categories to strengthen the model performance.

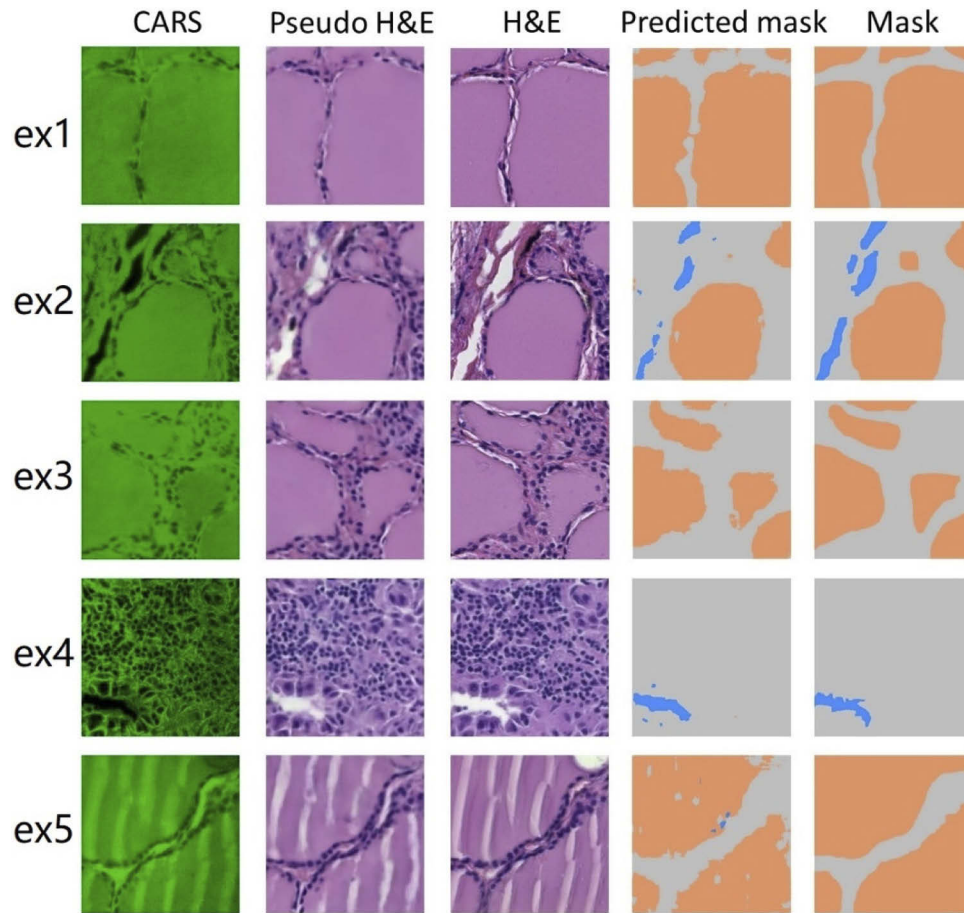


Fig. 5. Representative samples from the test results are showed. The shade corrected CARS images are inputs of UNet+/seg-cGAN. The columns of pseudo H&E and predicted mask are images generated by the model. Columns of label H&E and label mask images are gold standards of the real H&E stained images and segmentation masks of content categories.

In Sample 4 of Fig. 5, the black area on the bottom left of the CARS image was correctly classified and translated into pseudo H&E staining image, in spite of some missed details in the CARS image as compared to its real H&E staining image. Since the information cannot be invented, the non-existent tissue details in the CARS image will reasonably not appear in its pseudo H&E image.

To analyze how much UNet+/seg-cGAN has improved the translation, we compared UNet+ with other generators such as UNet and basic encoder-decoder without skip-connections. These generators with different filter numbers were also included in experiments. In a standard UNet generator, the filter number of the encoder block usually doubles compared to its previous block, while the filter number of the decoder block is its half compared to the previous one most of the time. The blocks in the same level in the encoder and decoder can connect each other by the skip

connection. We name the filter number of the first encoder block the basic filter number, which equals to the filter number of the final decoder block. Reasonably, the basic filter number should be the smallest block filter number in the generator.

Since there are two main improvements, including the generator upgrade from UNet to UNet+ and the simultaneous segmentation sub-model training, UNet cGAN, UNet+ cGAN and UNet+ cGAN with segmentation sub-model were trained with same basic filter numbers to illustrate the performance difference between different generators. A larger size UNet and UNet+ with 128 filter number are presented for reference.

Meanwhile, the UNet cGAN with segmentation sub-model was not constructed because the UNet+ has its special skip-connection layers, which are friendly to construct a segmentation sub-model, while the direct skip-connections between encoder blocks and decoder blocks in the standard UNet do not allow it to train a segmentation sub-model. Table 1 displays the results of models based on the test dataset. The first column of Table 1 includes the names of the models and followed by the basic filter numbers of their generators. The second column contains the numbers of generator parameters of each model. The parameter numbers are identical between the generators in UNet+ cGAN and UNet+ seg-cGAN with same basic filter number, because the sub-model training does not modify the generator networks. In the third column, the structural similarity index (SSIM) is a metric to measure how similar the pseudo H&E image is to its corresponding real H&E image.

Table 1. Performance of conditional GAN with different generators

Generators (basic filter number)	Number of parameters	SSIM	MSE
U-Net (64)	16,665,219	0.5148 (± 0.167)	0.0731 (± 0.025)
U-Net (128)	66,622,723	0.6911 (± 0.152)	0.0655 (± 0.014)
UNet+ (64)	21,948,483	0.7220 (± 0.145)	0.0296 (± 0.019)
UNet+ (128)	87,675,011	0.7591 (± 0.116)	0.0254 (± 0.011)
UNet+/seg (32)	5,501,987	0.6179 (± 0.125)	0.0129 (± 0.013)
UNet+ CycleGAN (64)	21,948,483	0.8125 (± 0.109)	0.0281 (± 0.089)
UNet+/seg (64)	21,948,483	0.8917 (± 0.071)	0.0079 (± 0.004)
UNet+/seg (128)	87,675,011	0.9066 (± 0.065)	0.0076 (± 0.003)

The images are all scaled between 0 and 1 before the SSIM score calculation following the same range rule for input arrays. The SSIM score is a real number between 0 and 1. The higher the SSIM score, the more similar the two images. In the last column is the mean square error (MSE), which measures the difference between two related images. The lower the MSE score, the more similar the two images.

Table 1 shows that the UNet+/seg strategy is the best performer with same or less parameters as measured by SSIM and MSE metrics. Note that because additional information from segmentation dataset was imported by training the CARS-H&E generator and CARS segmentation model simultaneously. The CARS segmentation model is a sub-model of CARS-H&E generator and shares its most parameters with the latter. The information about content categories helped the generator to better understand the CARS images and ameliorated the converging process. Since the objective problem is about translating from CARS images into H&E images, we do not need to train with the sub-model and directly obtained the UNet+ networks from UNet+ seg-cGAN, which has better performance than UNet+ from UNet+ cGAN. The generator's performance is improved while the generator structures stay identical. In addition, the performance of pseudo H&E stained images are slightly improved by using proposed stitching strategies. There are some prior works related to similar research [55,56]. In this paper, we compare our model to CycleGAN, even their inputs are based on 3 types of scans [41]. We re-trained a CycleGAN

model based on our data. The CycleGAN model performed worse than UNet+ seg-cGAN. UNet+ seg-cGAN recognizes better content categories, and therefore does less wrong color mapping. For example, in the ex2 of Fig. 6, the CycleGAN failed to recognize the background and did the wrong translation while UNet+ seg-cGAN did it correctly.

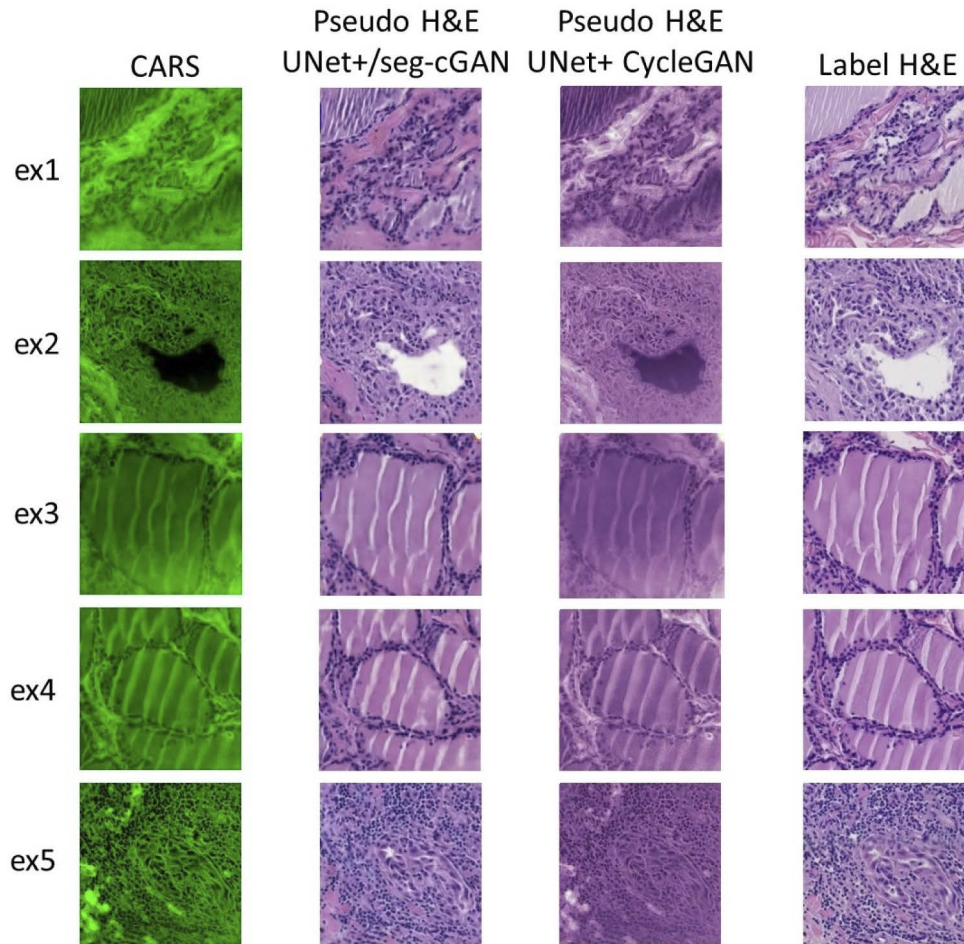


Fig. 6. Five test samples after stitching were shown above. The first column contains CARS images as UNet+/seg-cGAN's inputs. Pseudo H&E images generated from UNet+/seg-cGAN and UNet+ CycleGAN are located at the second column and third column. Key structures were merely kept comparing to the Real H&E images. However, UNet+ CycleGAN has much more color inversion problem than UNet+/seg-cGAN.

Table 2 shows the MSE and the SSIM of the stitching area between pseudo images and real H&E stained images. N/A is the control strategy (without any stitching method). The performance of the groups implemented with stitching are slightly better than the control. Figure 6 displays samples after stitching with the dynamic method. Zooming into the tissue details, there is almost no trace of splicing in pseudo H&E images. The tissue structures in the images become more coherent, and this reduces the probability of misdiagnosis.

Table 2. Stitching strategies' comparison.

Metrics	Mean	Median	Dynamic	N/A
MSE	0.0074 (± 0.006)	0.0071 (± 0.007)	0.0079 (± 0.004)	0.0085 (± 0.009)
SSIM	0.8810 (± 0.072)	0.8807 (± 0.074)	0.8917 (± 0.071)	0.8751 (± 0.079)

4. Discussion

In the model, UNet was replaced by UNet+ and the partial regularization strategy was applied with UNet+. The partial regularization allows UNet+/seg-cGAN to construct a sub-model simultaneously during training, which brings additional information into GAN's generator from CARS segmentation dataset and improves the converging process during training. Accordingly, the generator for CARS-H&E image translation has improved significantly without increasing its number of parameters.

The perceptual validation aims to test whether the pseudo H&E images generated by UNet+/seg-cGAN can fool humans. Just like most of the other GAN models, the output images are always slightly fuzzy. This consequently allows random participants to easily filter out the pseudo images with accuracy more than 70% without any pathological knowledge. Also, the outputs have slightly lower quality than the gold standards as the information that does not exist in CARS images but is present in targeted H&E images cannot be invented during image-to-image translation. The goal of our work is to generate H&E-like label-free CARS images, which are more helpful for pathologists in interpretation and help make the label-free modality more acceptable for pathologic diagnosis.

The pathologists participating in this study claimed that the results showed the huge potential of the model for the pathological diagnosis only based on translated images from CARS, however there are still some important structures missing during image translation. Figure 6 illustrates such examples. It appears that pseudo H&E images maintain most tissue architectural information and lots of cytological details. For example, in the top corner of ex1 of Fig. 6, the black artifacts may confuse pathologists during diagnosis. Also, the pseudo images may tend to have less quality than the actual H&E images due to the resolution of input CARS images, e.g., cellular clarity and nuclear or chromatin detail are less obvious.

The translation from CARS images will allow pathologists and clinicians to waive the tissue staining procedure during diagnosis to save the time and cost. The improvements made in this paper contributed to this ultimate goal. Coupling with endoscopy techniques, real-time acquisition and presentation of label-free CARS images in H&E format would realize pathology diagnosis *in vivo*.

Funding. John S. Dunn Foundation; Ting Tsung and Wei Fong Chao Family Foundation.

Acknowledgments. We also acknowledge Houston Methodist research pathology core for tissue staining and thank Rebecca Danforth for proofreading the manuscript.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. L. G. Rodriguez, S. J. Lockett, and G. R. Holtom, "Coherent anti-stokes Raman scattering microscopy: A biological review," *Cytometry, Part A* **69A**(8), 779–791 (2006).
2. L. Gao, H. Zhou, M. J. Thrall, F. Li, Y. Yang, Z. Wang, P. Luo, K. K. Wong, G. S. Palapattu, and S. T. Wong, "Label-free high-resolution imaging of prostate glands and cavernous nerves using coherent anti-Stokes Raman scattering microscopy," *Biomed. Opt. Express* **2**(4), 915–926 (2011).
3. C. L. Evans, X. Xu, S. Kesari, X. S. Xie, S. T. Wong, and G. S. Young, "Chemically-selective imaging of brain structures with CARS microscopy," *Opt. Express* **15**(19), 12076–12087 (2007).

4. Y. Yang, F. Li, L. Gao, Z. Wang, M. J. Thrall, S. S. Shen, K. K. Wong, and S. T. Wong, "Differential diagnosis of breast cancer using quantitative, label-free and molecular vibrational imaging," *Biomed. Opt. Express* **2**(8), 2160–2174 (2011).
5. L. Gao, Z. Wang, F. Li, A. A. Hammoudi, M. J. Thrall, P. T. Cagle, and S. T. Wong, "Differential diagnosis of lung carcinoma with coherent anti-Stokes Raman scattering imaging," *Arch. Pathol. Lab. Med.* **136**(12), 1502–1510 (2012).
6. O. Uckermann, R. Galli, S. Tamosaityte, E. Leipnitz, K. D. Geiger, G. Schackert, E. Koch, G. Steiner, and M. Kirsch, "Label-free delineation of brain tumors by coherent anti-Stokes Raman scattering microscopy in an orthotopic mouse model and human glioblastoma," *PLoS One* **9**(9), e107115 (2014).
7. R. Galli, V. Sablinskas, D. Dasevicius, A. Laurinavicius, F. Jankevicius, E. Koch, and G. Steiner, "Non-linear optical microscopy of kidney tumours," *J. Biophotonics* **7**(1-2), 23–27 (2014).
8. F. B. Legesse, A. Medyukhina, S. Heuke, and J. Popp, "Texture analysis and classification in coherent anti-Stokes Raman scattering (CARS) microscopy images for automated detection of skin cancer," *Comput. Med. Imag. Graph.* **43**, 36–43 (2015).
9. C.-H. Chien, W.-W. Chen, J.-T. Wu, and T.-C. Chang, "Label-free imaging of *Drosophila in vivo* by coherent anti-Stokes Raman scattering and two-photon excitation autofluorescence microscopy," *J. Biomed. Opt.* **16**(1), 016012 (2011).
10. D. Mayerich, M. J. Walsh, A. Kadjacsy-Balla, P. S. Ray, S. M. Hewitt, and R. Bhargava, "Stain-less staining for computed histopathology," *Technology* **03**(01), 27–31 (2015).
11. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydin, J. E. Zuckerman, T. Chong, and A. E. Sisk, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nat. Biomed. Eng.* **3**(6), 466–477 (2019).
12. J. J. Ojeda and M. Dittrich, "Fourier transform infrared spectroscopy for molecular analysis of microbial cells," in *Microbial Systems Biology* (Springer, 2012), pp. 187–211.
13. A. Alvarez-Ordóñez, D. Mouwen, M. Lopez, and M. Prieto, "Fourier transform infrared spectroscopy as a tool to characterize molecular composition and stress response in foodborne pathogenic bacteria," *J. Microbiol. Methods* **84**(3), 369–378 (2011).
14. N. Baletic, Z. Petrovic, I. Pendjer, and H. Malicevic, "Autofluorescent diagnostics in laryngeal pathology," *Euro. Archives Oto-Rhino-Laryngol. Head Neck* **261**(5), 233–237 (2004).
15. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," arXiv preprint arXiv:1406.2661 (2014).
16. J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," arXiv preprint arXiv:2001.06937 (2020).
17. X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.* **58**, 101552 (2019).
18. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134.
19. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
20. M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784 (2014).
21. Y. Lu, Y.-W. Tai, and C.-K. Tang, "Attribute-guided face generation using conditional cycleGAN," in *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 282–297.
22. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein GANs," arXiv preprint arXiv:1704.00028 (2017).
23. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning* (PMLR2017), pp. 214–223.
24. X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2794–2802.
25. X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "On the effectiveness of least squares generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(12), 2947–2960 (2019).
26. T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," arXiv preprint arXiv:1802.05957 (2018).
27. J. H. Lim and J. C. Ye, "Geometric GAN," arXiv preprint arXiv:1705.02894 (2017).
28. D. Tran, R. RanGANath, and D. M. Blei, "Hierarchical implicit models and likelihood-free variational inference," arXiv preprint arXiv:1702.08896 (2017).
29. Z. Gan, L. Chen, W. Wang, Y. Pu, Y. Zhang, H. Liu, C. Li, and L. Carin, "Triangle generative adversarial networks," arXiv preprint arXiv:1709.06548 (2017).
30. T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *International Conference on Machine Learning* (PMLR2017), pp. 1857–1865.
31. Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2849–2857.
32. E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," arXiv preprint arXiv:1506.05751 (2015).

33. Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2018), pp. 91–99.
34. R. Oulbacha and S. Kadoury, "MRI to CT Synthesis of the Lumbar Spine from a Pseudo-3D Cycle GAN," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (IEEE, 2020), pp. 1784–1787.
35. N. K. Singh and K. Raza, "Medical Image Generation Using Generative Adversarial Networks: A Review," *Health Informatics: A Computational Perspective in Healthcare*, 77–96 (2021).
36. R. Liu, Y. Lei, T. Wang, J. Zhou, J. Roper, L. Lin, M. W. McDonald, J. D. Bradley, W. J. Curran, and T. Liu, "Synthetic dual-energy CT for MRI-only based proton therapy treatment planning using label-GAN," *Phys. Med. Biol.* **66**(6), 065014 (2021).
37. X. Tie, S. K. Lam, Y. Zhang, K. H. Lee, K. H. Au, and J. Cai, "Pseudo-CT generation from multi-parametric MRI using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients," *Med. Phys.* **47**(4), 1750–1762 (2020).
38. H. Emami, M. Dong, S. Nejad-Davarani, and C. Glide-Hurst, "SA-GAN: Structure-Aware Generative Adversarial Network for Shape-Preserving Synthetic CT Generation," arXiv preprint arXiv:2105.07044 (2021).
39. J. K. Chan, "The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology," *Int. J. Surg. Pathol.* **22**(1), 12–32 (2014).
40. T. W. Bocklitz, F. S. Salah, N. Vogler, S. Heuke, O. Chernavskaya, C. Schmidt, M. J. Waldner, F. R. Greten, R. Bräuer, and M. Schmitt, "Pseudo-HE images derived from CARS/TPEF/SHG multimodal imaging in combination with Raman-spectroscopy as a pathological screening tool," *BMC cancer* **16**(1), 534 (2016).
41. P. Pradhan, T. Meyer, M. Vieth, A. Stallmach, M. Waldner, M. Schmitt, J. Popp, and T. Bocklitz, "Computational tissue staining of non-linear multimodal imaging using supervised and unsupervised deep learning," *Biomed. Opt. Express* **12**(4), 2280–2298 (2021).
42. J. Y. Lee, S.-H. Kim, D. W. Moon, and E. S. Lee, "Three-color multiplex CARS for fast imaging and microspectroscopy in the entire CHn stretching vibrational region," *Opt. Express* **17**(25), 22281–22295 (2009).
43. T. Peng, K. Thorn, T. Schroeder, L. Wang, F. J. Theis, C. Marr, and N. Navab, "A BaSiC tool for background and shading correction of optical microscopy images," *Nat. Commun.* **8**(1), 1–7 (2017).
44. L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 2414–2423.
45. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision* (Springer, 2016), pp. 694–711.
46. Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "IEEE Trans. Visual. Comput. Graphics," *IEEE Trans. Visual. Comput. Graphics* **26**(11), 3365–3385 (2020).
47. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vis. computing* **21**(11), 977–1000 (2003).
48. S. Nag, "Image registration techniques: a survey," arXiv preprint arXiv:1712.07540 (2017).
49. K. Murphy, B. van Ginneken, S. Klein, M. Staring, B. J. de Hoop, M. A. Viergever, and J. P. Pluim, "Semi-automatic construction of reference standards for evaluation of image registration," *Med. Image Anal.* **15**(1), 71–84 (2011).
50. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence* (2017).
51. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015), pp. 234–241.
52. H. Wang, Y. Li, and Z. Luo, "An Improved Breast Cancer Nuclei Segmentation Method Based on UNet++," in *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence* (2020), pp. 193–197.
53. M. Lei, J. Li, M. Li, L. Zou, and H. Yu, "An Improved UNet++ Model for Congestive Heart Failure Diagnosis Using Short-Term RR Intervals," *Diagnostics* **11**(3), 534 (2021).
54. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imaging* **39**(6), 1856–1867 (2020).
55. Y. Rivenson and A. Ozcan, "Toward a Thinking Microscope: Deep Learning in Optical Microscopy and Image Reconstruction," ArXiv abs/1805.08970 (2018).
56. B.-C. Chen, J. Sung, and S.-H. Lim, "Chemical imaging with frequency modulation coherent anti-Stokes Raman scattering microscopy at the vibrational fingerprint region," *J. Phys. Chem. B* **114**(50), 16871–16880 (2010).